Dixit, Avinash K.; Weibull, Jörgen W.

Working Paper

Political polarization

SSE/EFI Working Paper Series in Economics and Finance, No. 655

Political Polarization*

Avinash K. Dixit
Department of Economics
Princeton University

Jörgen W. Weibull
Department of Economics
Stockholm School of Economics

SSE/EFI Workning Paper Series in Economics and Finance No. 655

August 22, 2006. This revision: April 19, 2007.

Abstract

Failures of government policies often provoke opposite reactions from citizens; some call for a reversal of the policy while others favor its continuation in stronger form. We offer an explanation of such polarization, based on a natural bimodality of preferences in political and economic contexts, and consistent with Bayesian rationality.

Key words: polarization, voting, distinct priors, information.
JEL codes: D72, D74, D81, D82.

*We thank Cedric Argenton, James Fearon, John Londregan, Robert Powell, Jean Tirole, and Mark Voorneveld for helpful advice and comments on earlier versions of this paper. Dixit thanks the U.S. National Science Foundation, and both authors thank the Knut and Alice Wallenberg Research Foundation for financial support.

The failure of many economic policies, and indeed of other social policies or military actions, often invokes opposite reactions from different segments of the citizenry. Some argue that the failure indicates the need for a reversal of the policy, while others interpret the same evidence as showing that the policy was not followed fully or firmly enough and arguing for a strengthening or more zealous enforcement of existing measures.

A few examples will make the point. When the British economy was performing poorly under the Labour governments in the 1970s, the Conservatives led by Margaret Thatcher called for drastic market-oriented reforms, while traditional Labour supporters said that the real problem was the failure to adopt true Socialism. Similar divisions arose in former Socialist economies as their initial attempts at market reforms met with limited and mixed success, or in some cases outright failure and decline of the economies. One part of their populations wanted the reforms speeded up and made more drastic, while others wanted to slow down or even reverse the reforms and go back to many of the old Communist policies. In these instances, many individuals wanted broadly the same results – more output and growth – and observed the same outcomes of the prevailing policies – success or failure in various respects – but drew divergent conclusions from these observations. Similarly large splits of public opinion opened up in the United States in the late 1960s and early 1970s about the Vietnam war – whether to pursue it with greater force or to withdraw. Most of the opponents of the war agreed with the proponents that a democratic Vietnam would be desirable, but drew different inferences from the same events. Today we witness political polarization in many parts of the world on issues of discrimination, multiculturalism, religion, immigration, human rights, terrorism, civil war and nuclear armament.

How can we explain such increasing polarization of opinion even when both sides are broadly agreed on the objective of the policy and are observing the same evidence? One could simply appeal to biases in perception and reasoning, but it would be desirable to understand whether increased polarization is compatible with standard theories of statistical inference. We argue that the standard locational model of policy preferences that is used in political science and economics implies a natural bimodality that can generate temporary polarization under Bayesian updating.[1]

[1] In a working paper that appeared after our first draft was completed, Acemoglu, Chernozhukov and Yildiz (October 2006) develop a model that can even generate permanent divergence of beliefs. They provide a general theory for distinct priors, establish asymptotic results and consider applications to coordination games, asset trading and bargaining. While they require a positive prior probability that the observed signal is uninformative, our model of temporary polarization works even with surely informative signals.

1 Definition and General Propositions

We develop our argument in a simple model where the outcome of policy depends on some underlying unobservable "state of the world." Denote states of the world by points s in a set S of possible states. Each state of the world could be a fact about or aspect of how the world works; indeed it could be an entire "possible world". In the case of monetary policy, for example, the two states can be "Keynesian" and "monetarist" worlds.[2] The state of the world is fixed and unaffected by policy, but different policies may result in different outcomes.[3] Actual policy outcomes are often multidimensional, complex and not observed with much detail or accuracy. What can usually be observed is a summary indicator of the outcome, which is subject to random disturbances of measurement and estimation. For example, outcomes of monetary and fiscal policies include the effects on incomes and prices faced by millions of consumers and firms; what we observe is an index of inflation or unemployment constructed by the relevant bureau of statistics. Sometimes citizens are aware of only a binary indicator, such as an estimate of whether a policy is judged a success or failure. Denote the random disturbance affecting the observed magnitudes by $u \in U$. The product set $\Omega = S \times U$ is then our sample space over which probabilities are defined, with typical element (sample point) denoted by $\omega = (s, u)$.[4]

Neither the true state of the world s nor the disturbance u is observable. However, each individual can observe the actual policy x that is being pursued and some (potentially noisy) indicator y of the policy outcome, which in turn depends on the policy and the true state of the world. Thus there is a known functional relationship, $y = Y(x, s, u)$. Each individual has his or her own prior probability distribution over S, the possible states of the world. These priors may be thought of as initial "world views" or beliefs about the "true nature" of the world we live in. Upon observing y, the individual updates his or her prior, using Bayes' rule to obtain a posterior concerning the state of the world. This is the individual's revised (or confirmed) world view or belief.[5] The individual cannot in general infer s from x and y,

[2] Piketty (1995) considers a model of fiscal policy where individuals' priors about the extent of equality of opportunity in society and the effectiveness of individual effort differ in just such a way.

[3] The subsequent analysis can be generalized to dynamic situations in which the state of the world is not fixed but changes over time in part depending on policy.

[4] Even more generally, we could consider an abstract sample space Ω endowed with a sigma algebra $\mathcal{M}$ and ($\mathcal{M}$-measurable) random variables s and u that map sample points $\omega \in \Omega$ to states of the world, $s(\omega) \in S$, and errors, $u(\omega) \in U$, where U and S are sets in Euclidean spaces, see e.g. Billingsley (1999, pp. 16–42).

[5] Note that this whole situation is the opposite of the one of "agreeing to disagree" familiar to many economists (Aumann, 1974, see also Geanakoplos and Polemarchakis, 1982). There, individuals have common

not only because of the random disturbance u but also because the mapping $y = Y(x, s, u)$ from s to y, given x and u, need not be one-to-one. We will assume that the probability distribution of the random disturbance u is known, so no Bayesian revision of its probability distribution need be made, but even this can be generalized. We assume the random draws of s and u to be statistically independent.

Our concept of polarization of different individuals' probabilistic beliefs is illustrated in Figures 1 and 2, where the horizontal axis represents a one-dimensional spectrum of states s of the world. Figure 1 shows the prior and posterior cumulative distributions F, and Figure 2 shows the corresponding probability density functions f, for two individuals identified by the colors red and blue. The priors are shown as solid curves and the posteriors as dashed curves. The prior of red is to the left of that of blue in the sense of first-order stochastic dominance. The posterior of red is even farther to the left than her prior, and the posterior of blue is even farther to the right than his prior. We call this polarization, and examine when it can arise.

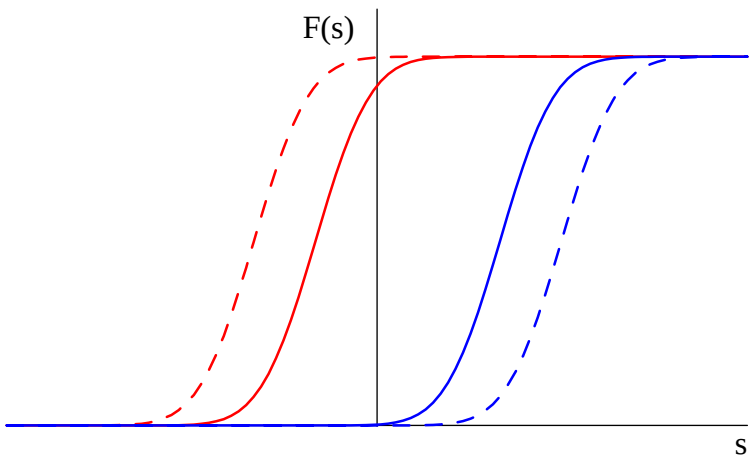

Figure 1: Polarization illustrated using cumulative distribution functions.

priors but get different observations; here, they have different priors but get common observations. Economic and game-theoretic analyses often rely on an assumption of common priors on the argument that it "enables one to zero in on purely informational issues" (Aumann 1976, p. 14) but recently departures from this assumption have been made to "zero in on open disagreement issues"; see Van den Steen (2001, p. 5) and Morris (1995). Indeed, open disagreement issues are often the essence of political problems.

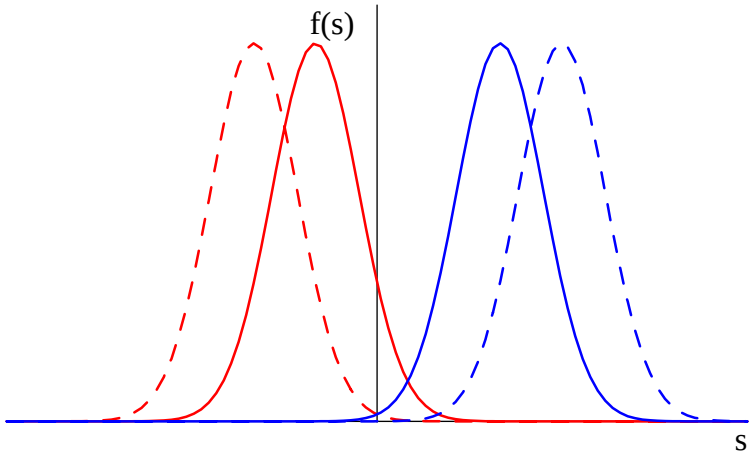

Figure 2: Polarization illustrated using density functions.

It helps to start with a condition that rules out such polarization. Let $g(y|x, s)$ be the probability density of the observable y conditioned on the policy x and state s. Polarization will be ruled out if, in response to a higher y, the posterior for s, calculated using Bayes' Rule, always shifts in the same direction in the sense of first-order stochastic dominance. It is well known that a necessary and sufficient condition for this is that $g(y|x, s)$ has the monotone likelihood ratio property (MLRP), namely that the likelihood ratio $g(y|x, s_1)/g(y|x, s_2)$ for any policy x and states $s_1 > s_2$ is a monotone (increasing or decreasing) function of y.[6]

MLRP holds for many standard distributions such as the normal, which is often used and has affected thinking about polarization. Suppose the observable y is a linear function of its arguments, with coefficients set equal to 1 by choice of units:

$$y = x + s + u.$$

Let an observer's prior distribution of s — his or her "initial world view" — be normal with mean μ and precision τ_s (reciprocal of variance), and let the known distribution of u be normal with mean 0 and precision τ_u. When x and y are observed and the distribution of s is updated according to Bayes' Rule, the posterior distribution for the state of the world (the observer's "revised world view") is normal with mean ν given by

$$\nu = (1 - \beta)\,\mu + \beta(y - x),$$

[6] This property was first identified by Rubin in the 1950's, further developed in Karlin and Rubin (1956) and introduced into the economics literature by Milgrom (1981).

where

$$\beta = \frac{\tau_u}{\tau_s + \tau_u},$$

a convex combination of the prior μ and the "signal" $y - x$,which we abbreviate as y'. The precision of the posterior is $\tau_s + \tau_u$. Thus in this model, the posterior is more precise than the prior, and a higher (lower) realization of the observable indicator y leads to a higher (lower) mean of s under the posterior distribution, for any given prior μ.

Suppose two observers have different prior means, say μ_1 and μ_2, and possibly different βs, β_1 and β_2. Choose the labels so that $\mu_1 > \mu_2$. When the two observe the same y and update their priors, let the posterior means be ν_1 and ν_2. Using the above formula for the posterior mean, we have the following three cases:

[1] If $y' > \mu_1$, then $\mu_1 < \nu_1 < y'$ and $\mu_2 < \nu_2 < y'$, so both distributions shift to the right.

[2] If $y' < \mu_2$, then $y' < \nu_2 < \mu_2$ and $y' < \nu_1 < \mu_1$, so both distributions shift to the left.

[3] If $\mu_2 < y' < \mu_1$, then $\mu_2 < \nu_2 < y'$ and $y' < \nu_1 < \mu_1$, so the two distributions shift toward each other.

In no case can the distributions shift as in Figures 1 and 2; polarization cannot occur. Users of the linear-normal model then have to explain the instances of belief polarization by invoking some biases of perception or learning to modify Bayesian updating. Gerber and Green (1999) review this literature.

We now illustrate how the monotone likelihood ratio property fails in some politico-economic situations where policies form a one-dimensional spectrum, such as left to right or dove to hawk. Think of a policy as a real number. In the example of monetary policy, this can be the rate of growth of the money supply, or the federal funds rate. If policy x is used when the state is s, this will generate a loss, $L(x, s)$. The individual knows the actual policy, x, and observes an indicator of the ensuing loss. In order to contrast our approach with the usual approach in current political science and economics, where individuals hold the same prior but have distinct preferences and hence loss functions, we here assume that everyone is agreed about the function L; disagreement is limited to probabilities of the states of the world. We do not deny the realistic possibility that people also have distinct preferences. However, heterogeneity of beliefs about the nature of the world we live in is also realistic, and here we focus on its consequences by leaving out the other possibility. Preference differences can be an additional reason for polarization.

An optimal policy $x^*(s)$ in state s is one that minimizes the loss $L(x, s)$. For the sake of simplicity we assume that the optimal policy is unique in every state. Thus everyone is agreed about the function $x^* : S \to \mathbb{R}$ that maps each state to its optimal policy, but people may disagree about what policy should be chosen in order to minimize the *expected* loss if they have distinct probabilistic beliefs about the true state of the world. For notational convenience, suppose that $x^*(s) \equiv s$.

The loss is not directly observed, but only an indicator $Y(x, s, u)$ is observed. We consider two cases of this.

Suppose, first, that the function Y only takes two values; either 0, "success," or 1, "failure" and let $X = S$ be a subset of the real line. Failure becomes more likely both when the policy is farther to the right of its optimum and when it is farther to the left of the optimum. Conversely, success becomes more likely when policy gets closer to its optimum from either direction. The individual knows the actual policy, x, and observes whether it succeeds or fails. So here y simply is "success or failure" of policy x. For $y \in \{0, 1\}$, let $g(y|x, s)$ denote the conditional probability of the event $Y(x, s, u) = y$, given the policy x, state s, and observation error u. For $y = 1$ ("failure"), this probability increases with the distance $|x - s|$ between the actual and optimal policy, say $g(1|x, s) = \varepsilon + \delta |x - s|$ for $\varepsilon, \delta > 0$ such that $g(1|x, s)$ is always less than one over the ranges of the variables in the context. Consider any two states s_1 and $s_2 > s_1$. Then the likelihood ratio for failure is

$$\frac{g(1|x, s_1)}{g(1|x, s_2)} = \frac{\varepsilon + \delta |x - s_1|}{\varepsilon + \delta |x - s_2|}$$

and that for success ($y = 0$) is

$$\frac{g(0|x, s_1)}{g(0|x, s_2)} = \frac{1 - \varepsilon - \delta |x - s_1|}{1 - \varepsilon - \delta |x - s_2|}.$$

For the MLRP to hold, one of these likelihood ratios, say the first, should always exceed the other. This is equivalent with the requirement that either $x > (s_1 + s_2)/2$ or $x < (s_1 + s_2)/2$ for all x and s. This not being the case, the posterior may just as easily shift in one direction as in the other, depending on the prior. Since all individuals know the policy and make the same observation as to its success or failure, their posteriors may shift in different directions depending on their priors concerning the state of the world.

Secondly, consider an example where a suboptimal policy generates a loss on a continuum scale. Common forms in the political science and economics literatures for the loss function

are

$$L(x,s) = [x - x^*(s)]^2 \qquad \text{and} \qquad L(x,s) = |x - x^*(s)|,$$

where x is the actual policy and $x^*(s)$ the optimal policy in state s. Suppose that the loss is observed only with noise: $y = Y(x,s,u) = L(x,s) + u$, where u is an observation error. Suppose the actual policy x is known, but $x^*(s)$ and u are not known or observed. Then the observer will infer that, conditional on y and x, the optimal policy in the current state satisfies

$$x^* = x \pm \sqrt{y-u} \qquad \text{and} \qquad x^* = x \pm (y-u)$$

in the respective cases. Whether the plus or the minus part gets more posterior probability weight, after x and y have been observed, and therefore whether the person's preferred policy shifts to the left or the right, depends on the person's prior. Thus the usual locational or spatial spectrum model of policy creates an automatic bimodality in revisions of beliefs about the optimal policy.

Here are some concrete examples of bimodality. [1] Suppose country A intervenes militarily in country B, and the result is violence and ethnic conflict in B. This could be happening either because each ethnic group in B supports its own militants to resist A's forces and the militants then turn on each other, or because A's forces are not strong enough to maintain law and order. That is, the ideal policy could be either no intervention or a much stronger intervention, and the actual policy may be failing in the respective cases because it is too much or because it is too little. [2] Suppose a country is experiencing high unemployment. Those who take a Keynesian view of the world may think this is because monetary policy is too tight, whereas those who take a monetarist view may think that the policy is too loose, and that businesses are not hiring because they think that the loose policy will lead to inflation and then to much higher interest rates.

The kind of polarization we find does not last for ever. Under mild technical conditions, the difference between the posteriors eventually goes to zero when these are successively updated following a sequence of observations (Blackwell and Dubins, 1962).[7] But there is no general guarantee that the convergence occurs monotonically. Our examples show how divergences can temporarily increase, and thereby help us understand the process of polarization in greater detail. Moreover, these examples show that political polarization can arise quite naturally and consistently with Bayesian updating, without any need to invoke

[7] See, however, Acemoglu et al (2006) who show that if the conditional success probabilities (our table at the bottom of p. 6) are unknown and individuals' subjective beliefs about these are sufficiently diffuse, their posteriors will not converge even asymptotically.

selective perception or biased learning. Of course these things exist in reality and can further aggravate the polarization.

2 Binary indicator of success

We now develop these two examples of polarization of individual prior beliefs into fuller models of political polarization, where the policy is chosen by majority rule applied to the votes of the same diverse individuals. Our modeling of politics is admittedly special, but does bring out some useful intuitions. Specifically, we assume that all voters vote, and that they vote sincerely, that is, each voter in each election votes for his or her currently most preferred alternative according to his or her current belief. We do not consider more sophisticated strategic, forward-looking behavior. The context we have in mind is that of an election with numerous voters, where each has a negligible probability of being pivotal to the outcome, and therefore no one has the ability to manipulate the outcome strategically.[8]

Suppose there are five states of the world, $s \in S = \{1, 2, 3, 4, 5\}$, and equally many voter types, $\theta \in \Theta = \{1, 2, 3, 4, 5\}$, with equally many voters of each type. Each voter type holds a distinct subjective prior about the true state of the world. Table 1 shows the prior probabilities at the start of period 1 with voter types as rows and states of the world as columns.

Table 1 - Prior probabilities for different types of voters

	s =1	2	3	4	5
θ =1	0.700	0.297	0.001	0.001	0.001
2	0.200	0.600	0.198	0.001	0.001
3	0.001	0.269	0.500	0.229	0.001
4	0.001	0.001	0.198	0.600	0.200
5	0.001	0.001	0.001	0.297	0.700

The general motivation behind these specific numbers is as follows. [1] Each type of voter

[8] See Laslier and Weibull (2007) for a rigorous analysis of this issue. Strategic voting under a common prior but private signals is modeled by Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1998). Experimental evidence for small electorates (3 or 6 voters) gives some (but not strong) support to strategic voting; see Guarnaschelli, McKElvey and Palfrey (2000). Degan and Merlo (2007) find that "by and large sincere voting can explain virtually all of the individual-level observations on voting behavior in presidential and congressional U.S. elections in the data."

assigns high positive probability to the state of the world corresponding to his type, a significant probability for it being one position away from his type, but very small probabilities for it being farther away.[9] [2] Extremist individuals attach higher probabilities to their own type being the "right" type. [3] The voter of type 3 has a slightly higher prior probability of the state being 2 than it being 4.

There are five policies: Far Left (FL), Left (L), Center (C), Right (R) and Far Right (FR), which we label $x = 1, 2, ..., 5$. Formally, $X = S$. The loss function equals 1 if the policy fails and 0 if it succeeds. Therefore minimization of expected loss is equivalent to minimizing the probability of failure. Then $x = 1$ (FL) is the (unique) optimal policy in state 1, $x = 2$ (L) the optimal policy in state 2 etc. Formally: $x^*(s) = s$ for all $s \in S$. We develop the example assuming that the true state of the world is 4, so $x = 4$ (R) is the optimal policy.

At any time, the voter can observe what policy is actually being followed, and can observe a binary indicator of the outcome (success or failure). Table 2 shows the probabilities of failure for each policy in each state of the world (with policies as rows and states of the world as columns):

Table 2 - Probabilities of failure for available policies in each state of the world

	s =1	2	3	4	5
x =1	0.01	0.2	0.4	0.7	1
2	0.2	0.01	0.2	0.4	0.7
3	0.4	0.2	0.01	0.2	0.4
4	0.7	0.4	0.2	0.01	0.2
5	1	0.7	0.4	0.2	0.01

The policy $x = 1$ is optimal in state 1 and fails with probability only 0.01, but it is farther from optimal in states 2, 3, 4, and 5, and fails with higher probabilities. Similarly for other policies. Thus there is a small probability that even the optimal policy will fail,[10] and failure probabilities rise when the actual policy is farther away from the optimal. Note

[9] Even a voter of one extreme type attaches positive (albeit very small) probability to the state of the opposite extreme type; the priors are non-dogmatic. This is to alleviate any concern that zero prior probabilities might be driving our results.

[10] As with our previous assumption of non-dogmatic priors, we introduce these small probabilities of failure of ideal policies so as to mitigate any reader's concern that our results are being driven by Bayesian updating on zero probability events.

that policy 3 is optimal in state 3, its failure probability is higher when the true state is 2 or 4, and higher still when the true state is 1 or 5. This is the natural bimodality that arises in the spatial model, and drives our results.

We assume that each type of voter knows the table of failure rates. In period 1, given their priors, each type's sincere vote is for the policy that coincides with his type. Preferences are single-peaked, and the outcome is the median voter's preferred policy, which in this case is 3. So policy $x = 3$ is adopted. Suppose, however, that policy 3 leads to failure. The voters now revise their priors using Bayes' Rule. The posterior from period 1 become the priors at the start of period 2. For each voter type, the posterior probability that the true state is j is proportional to

$$\Pr(\text{Failure} \mid s = j \text{ and } x = 3) * \text{Prior}\Pr(s = j).$$

The actual probabilities are found by normalizing these products. Table 3 shows these products, with voter types as rows and states as columns:

Table 3 - Products of prior and conditional probabilities

	s =1	2	3	4	5
θ =1	0.28	0.0594	0.00001	0.0002	0.0004
2	0.08	0.12	0.00198	0.0002	0.0004
3	0.0004	0.0538	0.005	0.0458	0.0004
4	0.0004	0.0002	0.00198	0.12	0.08
5	0.0004	0.0002	0.00001	0.0594	0.28

Then Table 4 shows (to four significant digits) the resulting posteriors, found by dividing each entry by the sum of all the cells in its row (still with voter types as rows and states as columns).

Table 4 - Period 1 posterior probabilities for different types of voters

	s =1	2	3	4	5
θ =1	0.8235	0.1747	0.0000	0.0006	0.0012
2	0.3949	0.5924	0.0098	0.0010	0.0020
3	0.0038	0.5104	0.0474	0.4345	0.0038
4	0.0020	0.0010	0.0098	0.5924	0.3949
5	0.0012	0.0006	0.0000	0.1747	0.8235

Now consider period 2 voting, given these as the new priors. Voter types 1, 2, 4, and 5 obviously vote for their corresponding policies 1,2, 4 and 5, respectively. But voter type 3's prior has become bimodal because the observed failure of policy 3 causes him to revise that probability drastically. So we must be more careful and calculate voter type 3's estimate of the probability of failure for all five policies in order to determine his most preferred policy, given his new prior. The failure probability of each policy $i \in X$ is

$$\Sigma_j \Pr(\text{Failure}|s = j \text{ and } x = i) * \text{New Prior} \Pr(s = j).$$

Table 5 shows the result of this calculation:

Table 5 - Voter type 3's assessment of policy failure probabilities

x =1	0.4291
2	0.1918
3	0.1925
4	0.2222
5	0.4670

So voter 3's preferences are still single-peaked, and his best choice is 2; this is his new most preferred policy. Therefore the median voter is at 2, and policy 2 is adopted – policy shifts in the "wrong" direction. Suppose policy $x = 2$ also leads to failure. Now the Bayesian revision yields posteriors shown in Table 6:

Table 6 - Period 2 posterior probabilities for different types of voters

	s =1	2	3	4	5
θ =1	0.9832	0.0104	3.5 E-5	0.0014	0.0049
2	0.8911	0.0668	0.0221	0.0045	0.0156
3	0.0040	0.0266	0.0495	0.9061	0.0138
4	0.0008	1.9 E-5	0.0038	0.4594	0.5360
5	0.0004	1.9 E-6	9.1 E-6	0.1081	0.8915

So voter type 3 has an epiphany – his probability distribution switches drastically to state 4. Hence, his most preferred policy is now 4. And the others become polarized: even type 4's most preferred policy now switches to 5.

This example serves to make three points: [1] Even slight asymmetries in initial beliefs can build into substantial differences. [2] Polarization can occur in a way that even voters

who are moderately biased in one direction come to favor the extreme policy in that direction. [3] The outcome of an election can be determined by the switching of a very small number of the centrist type 3 voters, but everyone else is polarized to favor extreme policies; therefore the outcome is likely to cause a lot of dispute and acrimony.

3 Observable continuous loss

Our second example allows policies to range over a continuum. Here we focus on the possibility that polarization can arise because the function $Y(x, s, u)$ is not monotonic in x; therefore we assume that there is no error term u.Voters are denoted by an index θ ranging over the unit interval, and for simplicity of exposition they are assumed to be uniformly distributed over this range. Each voter has a continuous prior distribution about the true state of the world s, where s ranges over $S = \mathbb{R}$, the real line. The prior probability density function of voter θ is f_θ, that is, θ assigns probability $f_\theta(s)\,ds$ to the event that the true state of the world lies in the interval $(s, s + ds)$. Let $X = S$.

The loss associated with an outcome is equal to the absolute value of the difference between the actual policy and state. The optimal policy in any state s is thus $x = s$, that is, $x^*(s) \equiv s$. Writing x for the actual policy, the observable is therefore given by

$$y = L(x, s) = |x - s|.$$

Voter θ likes best the policy x that minimizes the expected loss, $\mathbb{E}_\theta[L(x, s) \mid x]$, calculated under his prior. This is θ's most preferred policy, given his or her prior. We write this loss as:

$$c_\theta(x) = \int_{-\infty}^{+\infty} |x - s|\, f_\theta(s)ds = \int_{-\infty}^{x} (x - s)\, f_\theta(s)ds + \int_{x}^{+\infty} (s - x)\, f_\theta(s)ds.$$

The first-order condition for its minimization is

$$c'_\theta(x) = \int_{-\infty}^{x} f_\theta(s)ds - \int_{x}^{+\infty} f_\theta(s)ds = 2F_\theta(x) - 1 = 0,$$

where F_θ is the cumulative distribution function for voter θ's prior. And

$$c''_\theta(x) = 2F'_\theta(x) = 2f_\theta(x) > 0,$$

so the second-order condition is globally satisfied. Therefore the optimum is given by $F_\theta(x) =$

$\frac{1}{2}$. Hence, voter θ's most preferred policy x_θ is the median of his prior distribution over the states of the world. Moreover, each voter's preferences are single-peaked around his most preferred policy.

Consider the first election (Period 1) under this set-up. Under majority rule, the median of the most preferred policies becomes the chosen policy. To keep the notation simple, suppose this is the point 0 on the policy spectrum. Suppose the optimal policy, in the true state of the world, is different from this; for the sake of definiteness suppose the true state is $s = 1$ and hence $x^* = 1$.

The actual policy and the loss are by assumption observable without error. These observations enable people to infer that the true state must be either $s^+ = x + L(x, s)$ or $s^- = x - L(x, s)$. Thus the continuous prior is updated to a two-point posterior. To keep the notation simple again, suppose $L(x, s) = 1$. With $x = 0$, the posteriors then become concentrated on 1 and -1. From Bayes' rule, the posterior probabilities for voter θ are

$$\Pr[s = 1] = \frac{f_\theta(1)}{f_\theta(1) + f_\theta(-1)}$$

and $\Pr[s = -1] = 1 - \Pr[s = 1]$.

Suppose there is a number $z > \frac{1}{2}$ such that the priors of voters in the range $0 < \theta < z$ satisfy $f_\theta(-1) > f_\theta(1)$, and the priors of voters in the range $z < \theta < 1$ satisfy $f_\theta(-1) < f_\theta(1)$. Figure 3 below illustrates this, with red curves showing the prior densities of voters $\theta < z$ and the blue curves those of voters $\theta > z$.

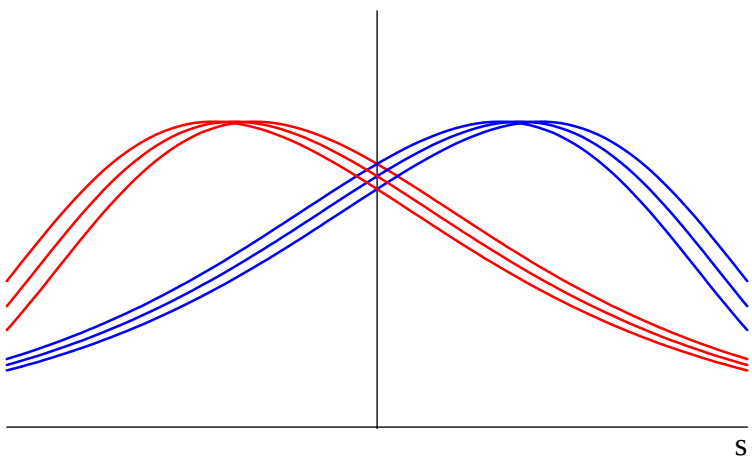

Figure 3: Prior densities for different types of voter.

Then the voters in $[0, z)$ have posteriors with $\Pr(s = -1) > \frac{1}{2} > \Pr(s = 1)$, and those in $[z, 1]$ have posteriors with $\Pr(s = -1) < \frac{1}{2} < \Pr(s = 1)$. These posterior become the new priors in the next election (Period 2). Therefore in that election more than half of the voters vote for the policy $x = -1$, their new most preferred policy, and fewer than half vote for the policy $x = 1$, their new most preferred policy. Whereas the election in Period 1 was a contest with a continuum of opinions leading to a moderate policy (albeit not the optimal, given the state), the election in Period 2 is polarized between two quite distinct positions, and the choice shifts away from the optimal policy. This can happen even if z is very close to one half.

With the optimal policy at 1 and the actual policy at -1, the outcome in Period 2 will be a loss equal to 2. With the priors concentrated on 1 and -1, and the actual policy at -1, this loss can arise only if the optimal policy is 1. Therefore Bayesian updating will lead to a convergence of opinions at the optimal policy, and that policy will be adopted unanimously in the election of Period 3.

In this example, once again quite small differences among voters can create polarization, and non-monotonic shifts in priors. However, the special structure with no error term in the loss function leads to quick reversal of the polarization and convergence to the optimal policy. An error term with a suitably large dispersion can slow down this process. We omit the details because the algebra gets complicated.

4 Concluding comments

We have seen how an electorate can become polarized and policies can shift away from the optimal, when the observable indicators of policy outcomes are not monotonic in the policy choice, and how such polarization is perfectly consistent with voters agreeing on values and using Bayesian updating and vote for the conditionally optimal policy given their information.

Political polarization entails quite serious risks; political debates get bitter and the very existence of a civil society may be threatened. Current examples are policies concerning discrimination, immigration, gender, religion, welfare state, human rights, terrorism, civil wars, national sovereignty and nuclear armament. One way to reduce these risks, therefore, is to attempt to create observable indicators that are not bimodal like the ones above, and satisfy the monotone likelihood ratio property. Of course that can still leave untouched the additional problems caused by biased perception and learning. Moreover, such indicators

may be hard to identify. However, our argument unambiguously supports the case for searching out and publicizing such indicators—under the here maintained hypothesis that people broadly agree on values but may have differing beliefs about the world. Contrary to the current tendency in many countries to avoid high-lighting socially and politically controversial and pressing issues, our simple examples suggest that political polarization may be reduced rather than increased if instead more information about the factual current situation and the effect of employed policies are made available in the public debate, even when the issues at hand are controversial.

References

Acemoglu, D., V. Chernozhukov and M. Yildiz. 2006. "Learning and Disagreement in an Uncertain World", Department of Economics, M.I.T., Working Paper 06-28.

Aumann, R. 1976. "Agreeing to Disagree." Annals of Statistics 4: 1236-1239.

Austen-Smith, D. and J. S. Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem", American Political Science Review 90:34-45.

Feddersen, T. and W. Pesendorfer. 1998. "Convicting the Innocent: the Inferiority of Unanimous Jury Verdicts", American Political Science Review 92:23-35.

Billingsley, P. 1999. Probability and Measure. New York: Wiley.

Blackwell, D. and L. Dubins. 1962. "Merging of Opinions with Increasing Information." Annals of Statistics 33: 882-886.

Degan, A. and A. Merlo. 2007. "Do Voters Vote Sincerely?" University of Pennsylvania, Penn Institute for Economic Research (PIER) Working Paper No. 07-006, available at http://ssrn.com/abstract=961827 .

Geanakoplos, J. and H. Polemarchakis. 1982. "We Can't Disagree Forever." Journal of Economic Theory 28: 192-200.

Gerber, A. and D. Green. 1999. "Misperceptions About Perceptual Bias." Annual Review of Political Science 2: 189-210.

Guarnaschelli S., R. McKelvey and T. Palfrey. 2000. "An Experimental Study of Jury Decision Rules", American Political Science Review 94: 407-423.

Karlin S. and H. Rubin. 1956. "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio", Annals of Mathematical Statistics 27: 272-299.

Laslier, J.-F. and J. Weibull. "Incentives for Informative Voting." Working paper, Ecole Polytechnique and Stockholm School of Economics, Feburary 2007.

Milgrom, P. 1981. "Good News and Bad News: Representation Theorems and Applications." Bell Journal of Economics 12: 380-391.

Morris, S. 1995. "The Common Prior Assumption in Economic Theory", Economics and Philosophy 11: 227-253.

Piketty T. 1995. "Social Mobility and Redistributive Politics", Quarterly Journal of Economics 110: 551-584.

Van den Steen E. 2001. "Essays on the Managerial Implications of Differing Priors", Ph D dissertation, Stanford Business School.